利用改进EfficientNetV2和无人机图像检测小麦倒伏类型

龙佳宁[1,2], 张 昭[1,2*], 刘晓航[1,2], 李云霞[1,2], 芮照钰[1,2], 余江帆[1,2], 张 漫[1,2], FLORES Paulo[3], 韩哲雄[4,5], 胡 灿[6], 王旭峰[6]

(1. 中国农业大学 信息与电气工程学院,北京 100080, 中国; 2. 中国农业大学农业农村部农业信息获取技术重点实验室,北京 100083,中国; 3. 北达科他州州立大学 农业与生物工程系,北达科他州法戈 58102,美国; 4. 韩国江原大学 生物系统工程系,江原道春川 24341,韩国; 5. 韩国江原大学 智慧农业交叉学科,江原道春川 24341,韩国; 6. 塔里木大学 机械电气化工程学院,新疆阿拉尔 843300,中国)

摘 要: [**目的/意义**] 不同类型的小麦倒伏(根部倒伏、茎部倒伏)对产量和质量会产生不同影响。本研究旨在通过无人机图像对小麦倒伏类型进行分类,并探究无人机飞行高度对分类性能的影响。[**方法**] 研究设置3个无人机飞行高度(15、45、91 m)来获取小麦试验田的图像,并利用自动分割算法生成不同高度的数据集,提出一种EfficientNetV2-C改进模型对其进行分类识别。模型通过引入CA(Coordinate Attention)注意力机制来提升网络特征提取能力,并结合CB-Focal Loss(Class - Balanced Focal Loss)来解决数据不均衡对模型分类准确度的影响。[**结果和讨论**] 改进的EfficientNetV2-C表现最佳,平均准确率达到93.58%。对比未改进的4种机器学习分类模型(支持向量机(Support Vector Machine,SVM)、K最近邻(K Nearest Neighbor,KNN)、决策树(Decision Tree,DT)和朴素贝叶斯(Naive Bayes,NB))与两种深度学习分类模型(ResNet101和EfficientNetV2),其中EfficientNetV2在各个高度下表现最优,平均准确率达到82.67%。无人机飞行高度对4种机器学习分类器性能无显著影响,但随飞行高度上升,由于图像特征信息损失,深度学习模型的分类性能下降。[**结论**] 改进的EfficientNetV2-C在小麦倒伏类型检测方面取得了较高的准确率,为小麦倒伏预警和农作物管理提供了新的解决方案。

关键词: 小麦倒伏类型;图像处理;深度学习;不平衡数据;机器学习;无人机

中图分类号: S513;TP391.41 **文献标志码:** A **文章编号:** SA202308010

引用格式:龙佳宁, 张昭, 刘晓航, 李云霞, 芮照钰, 余江帆, 张漫, FLORES Paulo, 韩哲雄, 胡灿, 王旭峰. 利用改进EfficientNetV2和无人机图像检测小麦倒伏类型[J]. 智慧农业(中英文), 2023, 5(3): 62-74. DOI: 10.12133/j.smartag.SA202308010

LONG Jianing, ZHANG Zhao, LIU Xiaohang, LI Yunxia, RUI Zhaoyu, YU Jiangfan, ZHANG Man, FLORES Paulo, HAN Zhexiong, HU Can, WANG Xufeng. Wheat lodging types detection based on UAV image using improved EfficientNetV2[J]. Smart Agriculture, 2023, 5(3): 62-74. DOI: 10.12133/j.smartag.SA202308010 (in Chinese with English abstract)

1 引 言

小麦作为世界最主要的农作物之一,其产量对国家粮食安全有着重要的影响。倒伏是影响小麦产量的主要原因[1,2]。作物的倒伏类型主要分为茎部倒伏与根部倒伏。茎部倒伏多由自然灾害引起;根部倒伏主要由土壤条件(如氮肥过多)、病虫害或耕作不当引起[3]。研究表明,小麦倒伏会造成减

收稿日期:2023-08-04

基金项目:国家重点研发计划项目(2022YFD2001500)

作者简介:龙佳宁,研究方向为农业机器人。E-mail:614020890@qq.com

*通信作者:张 昭,博士,教授,研究方向为农业智能装备。E-mail:zhaozhangcau@cau.edu.cn

copyright©2023 by the authors

产，增加收获难度，延长收获时间，并降低小麦品质[4-6]。其次，小麦的倒伏时间与类型是影响产量的主要原因之一，且根部倒伏的时间越早对小麦产量影响越大[7]。因此，对小麦倒伏类型的实时监测与评估对于小麦产量和品种选育至关重要。传统的小麦倒伏类型监测方法主要依靠人工现场评估，费时费力，并且由于人工评价的主观性，可能导致评估结果缺乏客观性和准确性[8]。随着计算机技术的发展，视觉监测手段作为一种低成本、准确、快速、客观的作物倒伏类型检测方法，对于小麦的实时监测和倒伏类型评估具有重要意义，能够为提高小麦产量和品种选育提供关键支持[9-11]。

近年来，随着无人机技术与传感器技术的快速发展，近地面遥感技术被广泛应用于农业生产领域。与卫星技术和三维图像技术相比，无人机搭载视觉传感器获取图像的方法因其高机动性、实时性和经济性在农业生产领域取得令人满意的结果[12, 13]。然而，考虑到应用的实时性与传感器成本等问题，性价比较高的RGB相机是检测小麦倒伏类型的最优选择。通过RGB相机采集的高分辨率图像进行倒伏类型的监测，需要开发适合的算法。现有针对图像处理的方法，根据特征提取的方法可分为传统机器学习和深度学习[14]。传统机器学习在提取特征时，通常需要手工选择或设计特征，随后将提取的特征输入到机器学习分类器，如使用支持向量机（Support Vector Machine，SVM）、随机森林（Random Forest，RF）和K最近邻（K Nearest Neighbor，KNN）进行小麦倒伏与非倒伏的分类[15]。Rajapaksa等[16]通过提取图像的灰度共生矩阵（Gray-level Co-occurrence Matrix，GLCM）训练SVM分类器，实现小麦与油菜的倒伏预测。同时，多项研究证明基于深度学习的分类器要优于机器学习分类器[14-17]。基于深度学习的分类模型在提取特征时，能够自动从原始数据中学习高低层次的抽象特征，而机器学习分类器需要手动提取特征。Yu等[18]提出一种结合小麦不同生长阶段无人机图像特征的倒伏检测模型。该模型以金字塔场景分析网络（Pyramid Scene Parseing Network，PSPNet）模型为基础，结合长短时记忆结构（Long Short-Term Memory，LSTM）、卷积注意力模块（Convolutional Block Attention Module，CBAM）和Tversky损失函数进行改进。改进后的Lstm_PSPNet（Long Short-term Memory Pyramid Scene Parseing Network）预测精度均高于90%，能够有效地预测不同生长阶段的倒伏区域。Zhang等[11]采用无人机提取小麦试验田图像，通过提取图像的5种特征，并采用3种机器学习分类模型RF、卷积神经网络（Convolutional Neural Network，CNN）和SVM对不同日期的倒伏小麦进行分类，结果显示，不同日期采集的数据并不影响模型的分类性能，且GoogLeNet的平均准确率最高为93%。但上述及现有研究中，大多将小麦倒伏状态粗略地划分为倒伏与非倒伏，并未进一步对倒伏类型细化，给小麦种植者选择抗倒伏类型品种与田间管理带来困难。Neupane等[19]利用无人机在3种不同高度（40、50、60 m）下获取香蕉的RGB图像，采用基于Inception-V2的Faster R-CNN作为主要特征提取器，检测图像中的香蕉并计数，结果表明，无人机的飞行高度会影响模型检测的性能与计数的准确性。然而，现有研究缺乏对无人机飞行高度是否影响分类模型性能的验证。

针对上述研究存在的问题，本研究基于无人机图像提出一种小麦试验田自动分割算法。这种算法能够根据需求将获取的图像自动分割并保存。随后提出一种改进的EfficientNetV2-C深度学习分类模型，引入增强特征提取能力的CA（Coordinate Attention）注意力机制与用于解决类别不平衡问题的损失函数CB-Focal Loss（Class-Balanced Focal Loss），并对比4种机器学习分类模型与深度学习模型分类性能，探讨了无人机飞行高度对模型分类性能的影响。

2 材料与方法

为检测生长过程中小麦的倒伏类型，并探究无人机飞行高度对分类模型的影响，研究按照如下流程进行（图1）。首先，无人机在3个不同高度（15、45、91 m）下采集小麦试验田的高清数字图像，并通过人工方式获取其倒伏类型。然后，通过设计的自动分割算法为每个高度生成单独的数据集。其次，对比传统机器学习与深度学习模型分类

性能，选择最优分类性能模型，并进一步改进。最后，对改进后的网络模型进行评价，以验证其在小麦倒伏分类问题上的效果。

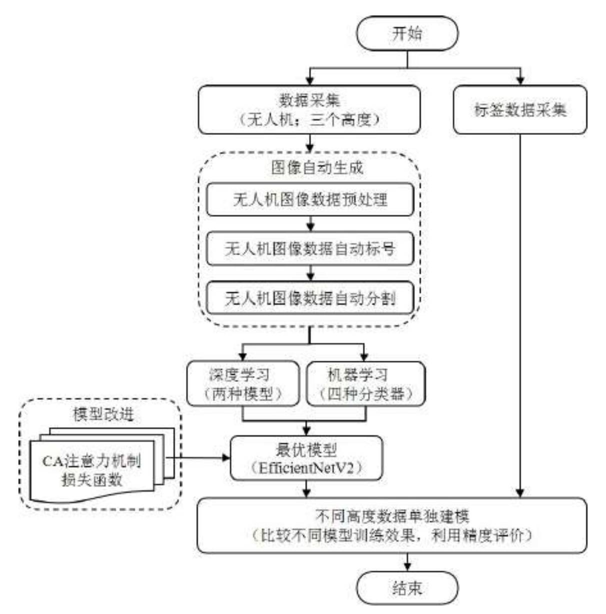


图1 小麦倒伏类型检测序流程图

Fig. 1 Flowchart of wheat lodging types detection

2.1 数据采集

在本研究中，使用大疆精灵Phantom 4 Pro V2.0无人机采集RGB图像，像素大小为5472×3078。拍摄时间为2020年8月23日，拍摄地点为美国北达科他州汤普森市（UTM WGS 84 14 N），当天光照条件充足。为保证能够完整拍摄小麦试验田，无人机飞行最低高度为15 m，最高高度为91 m，并选取45 m作为中间参考值。通过无人机在3个不同的飞行高度（15、45、91 m）采集小麦试验田的高清图像。其中，15 m时图像分辨率为0.4像素/cm；45 m时图像分辨率为1.2像素/cm；91 m时图像分辨率为2.5像素/cm。所有数据均来自同一天。为获得小麦试验田真实的倒伏类型信息，在无人机采集图像后，由农艺专家对试验田进行主观观察，并将麦田的倒伏类型分为未倒伏、根部倒伏和茎部倒伏，以此准确获取倒伏情况，为后续研究提供可靠的数据基础。

2.2 数据自动生成

本研究采用无人机采集的高清图像进行小麦试验田的分析，共包含464块小试验田。这些试验田分为两类：116块尺寸为1.5 m×15.0 m的矩形小麦试验田，以及348块尺寸为1.5 m×3.7 m的矩形小麦试验田。以往传统的试验田分割方法需要人工手动操作，耗费大量时间和精力。为解决这一问题，本研究提出一种自动分割算法，步骤如图2所示。该算法可以根据实际需求自动分割完整的小麦试验田。

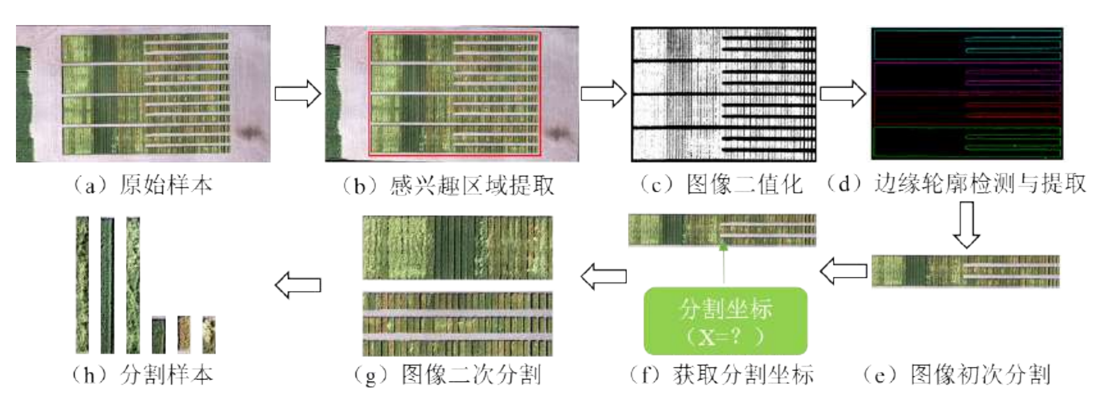


图2 模型数据集自动生成过程图

Fig. 2 Diagram of the process of automatic model dataset generation

原始试验田中包含两种长度的矩形试验田。为获取单个分割样本，本研究需对无人机获取的原始样本多次分割，以获取单个分割样本在图像中的结构参数与类型。鉴于小麦试验田在原始样本中所占区域较大，并且其颜色与土壤颜色相差较大，本研究选择超绿图像分割方法来提取感兴趣区域。具体操作步骤如下。

（1）感兴趣区域提取。将原始样本的RGB图像利用超绿算法提取绿色部分并二值化，随后利用边缘轮廓检测算法提取二值图的轮廓并根据轮廓大

小标号排序，最终选取最大轮廓并获取4个角点的坐标值根据实际图片进行微调，从而分离出主体与背景，如图2（b）所示，实现感兴趣区域的提取。

（2）图像初次分割。如图2（c）所示，将感兴趣区域二值化，其中白色像素（值=1）表示作物，黑色像素（值=0）表示土壤。随后，利用先腐蚀后膨胀的算法填充白色区域内细小的空洞，并连接邻近像素，平滑边界。最终，如图2（d）所示，采用Canny等算子提取边缘轮廓，以获取4个角点在图像中的像素坐标对图像进行初次裁剪，如图2（e）所示，并计算长矩形单个试验田样本的结构参数。

（3）分割坐标值自动获取与图像二次分割。如图2（f）所示，为获取短矩形试验田的结构参数，首先对初次分割的图像进行二值化，并提取边缘轮廓；其次通过遍历像素点并累加像素值的方法，寻找像素累加值突增点，从而获取长矩形与短矩形试验田的分割坐标点（X）的值；最终根据X值分割出长矩形与短矩形试验田，如图2（g）所示，并获取短矩形单个试验田结构参数。

（4）根据上述步骤获取的两种矩形结构参数，算法将会按照实际需求给小麦试验田自动标号，并将其分割成单独的样本，如图2（h）所示。最终，根据标号值顺序保存424张分割样本图像。

2.3 数据预处理

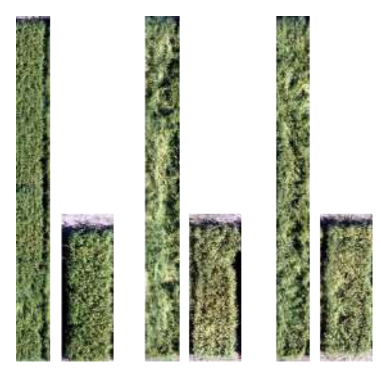
2.3.1 数据集分类

通过自动分割算法将无人机图像分为424张图像，结合农艺专家获取的倒伏类型数据，将无人机图像按图3标准，分为未倒伏（202张）、茎部倒伏（151张）和根部倒伏（71张）3类。

2.3.2 数据增强

本研究采集的数据量较小且样本类别分布不均匀。为增加模型训练的样本多样性，提升模型泛化能力，改善模型的鲁棒性，并减少过拟合的风险，对小麦倒伏数据进行数据增强处理。数据增强方式主要包括离线数据增强和在线数据增强两种方式。本研究采用适用于数据量较小的在线增强方式对数据进行增强操作，如图4所示，增强方法包括5种：

（1）缩放（Resize）操作。通过缩放操作将所有图像尺寸统一。



（a）未倒伏 （b）茎部倒伏 （c）根部倒伏

图3 不同小麦倒伏类型无人机图像示意图

Fig. 3 Schematic diagram of different wheat lodging types based on UAV images

（2）随机旋转（Random Rotation，RR）。根据概率将图像进行随机旋转，模拟在不同视角下拍摄的图像增加数据类型。

（3）水平、垂直镜像（Horizontal and Vertical Mirroring，HVM）。以X或Y轴为镜像轴，翻转图像，增加数据多样性。

（4）随机遮挡（Random Distribution，RD）。模拟真实环境视野遮挡。

（5）随机颜色抖动（Random Color Dithering，RCD）。更改图像的对比度、明亮度、颜色等模拟在不同光照条件下拍摄的图像。







（a）原图 （b）镜像 （c）旋转 （d）遮挡 （e）颜色抖动

图4 数据增强方式可视化图

Fig. 4 Visualization of the data enhancement approach

2.3.3 数据划分

本研究通过数据增强共获取2120张图像，将数据集以8：1：1的比例随机分成训练集、验证集和测试集。其中，训练集图像为1696张，验证集与测试集图像各212张。训练集和验证集用于模型训练，测试集用于模型验证。

2.4 小麦倒伏类型分类模型

2.4.1 机器学习分类模型

（1）支持向量机。SVM是一种常用于分类和回归分析的机器学习算法[20]。其核心思想是找到一个最优的超平面，以最大程度地分开不同类别的数据样本。SVM在数学模型结构构建中简单且有效，对小样本、非线性和高维度数据表现优异，广泛应用于图像信息提取、模式识别和数据挖掘等领域[21]。

（2）K最近邻（K-Nearest Neighbor，KNN）。作为机器学习算法中最基础的算法之一，简单易懂，并且对于非线性的数据集表现良好，因此常用于一些分类任务模型[22]。KNN的核心思想是通过计算待预测样本属性和标签样本属性之间的距离来进行分类。通过查看最近邻居所属的标签，采用多数表决法来判断样本所属类别，从而实现分类预测。

（3）决策树（Decision Tree，DT）。DT是一种监督学习模型。通过树形结构处理离散和连续特征，无需预处理。核心思想是通过决策规则对输入数据划分，以实现分类。从根节点开始，逐步划分样本数据为不同类别或数值。随后，在构建过程选择最佳特征划分数据，剪枝简化决策树，避免过拟合。最终，根据特征判断分类输入样本。

（4）朴素贝叶斯（Naive Bayes，NB）。该分类算法以概率统计知识为基础，核心思想是基于贝叶斯定理和特征之间条件独立性假设。根据图像提取颜色直方图特征，并利用训练数据统计每个类别出现的概率。随后，根据其出现的概率，计算条件概率。最后通过后验概率，将待分类样本分配给具有最有可能的类别，作为预测结果。

2.4.2 深度学习分类模型

基于深度学习CNN的特征提取方法被广泛应用于图像处理任务。其核心思想是通过多层神经网络学习高级抽象特征和表达。相较于传统的机器学习方法，深度学习模型能够自动从原始数据中学习到更加复杂的特征表达。本研究采用ResNet和EfficientNetV2作为小麦倒伏类型检测的基础模型。ResNet模型[23]引入了残差块，使网络架构能够有更深层的训练，有助于捕获更复杂和抽象的特征。由于输入网络的数据尺寸较大，本研究采用更深层和更多通道数的ResNet101，以更好地提取图像的特征用于麦倒伏类型检测。EfficientNetV2[24]遵循EfficientNet的思想，通过等比例地调整网络的深度、宽度和分辨率，在不增加计算复杂性的情况下提高模型性能。EfficientNetV2引入新的模块Fused-MBConv和渐进式学习策略来进一步改进模型性能。EfficientNetV2整体解决了因图像尺寸问题导致的训练速度慢和模型精度差等问题。

2.5 基于EfficientNetV2模型结构改进

本研究提出一种基于EfficientNetV2的改进模型，在不增加计算量的前提下，提升模型分类精度，并针对不平衡数据集能够得到较好的分类效果。图5显示了改进模型的整体框架。输入网络的数据为424张不同尺寸的小麦试验田分割图像。首先采用ConvBNAct普通卷积层改变输入数据的通道数。随后，通过神经网络结构搜索（Neural Architecture Search，NAS）优化训练效率和参数效率，模型前3层采用Fused-MBConv，后3层采用结合CA注意力机制改进的MBConv-C卷积层，通过连接输出层将数据输入给分类器。最终，采用解决难易样本数据不均衡的损失函数渐进式优化分类器，提升模型针对不均衡数据的分类效果。

2.5.1 基于CA注意力机制改进的MBConv-C

图像分类任务中卷积模块常用于提取图像特征。ResNet分类网络采用残差结构的卷积提取网络特征。但残差结构的卷积采用先降维后升维的方式提取网络特征，在操作过程中可能导致信息丢失等问题[25]。本研究采用倒残差结构的Fused-MBConv和MBConv-C模块，以先升维再降维形成了稀疏特征，降低了信息损失。其中MBConv-C模块首先将输入的特征通过包含BN和SiLU激活函数的普通卷积进行升维操作，并采用深度卷积减少运算量和参数数量，随后通过注意力机制分配通道权重，在接入包含BN层的普通卷积进行降维，最后接入Dropout层输出特征。通过MBConv-C模块使得网络参数和计算量大大降低。在卷积过程中注意力机制的选择将会极大影响模型分类的性能，未改进的MBConv-C采用SE（Squeeze and Excitation）注意力机

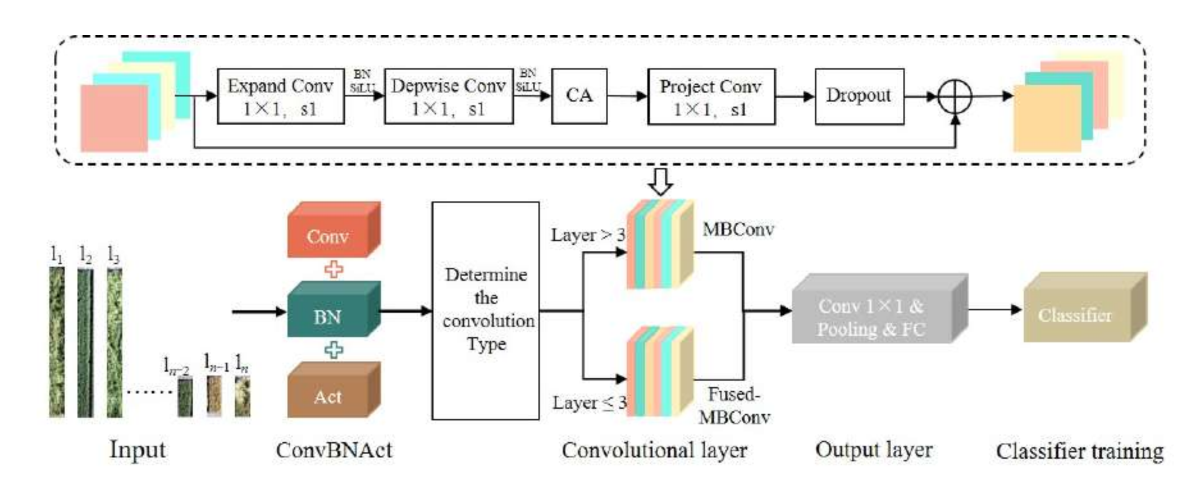
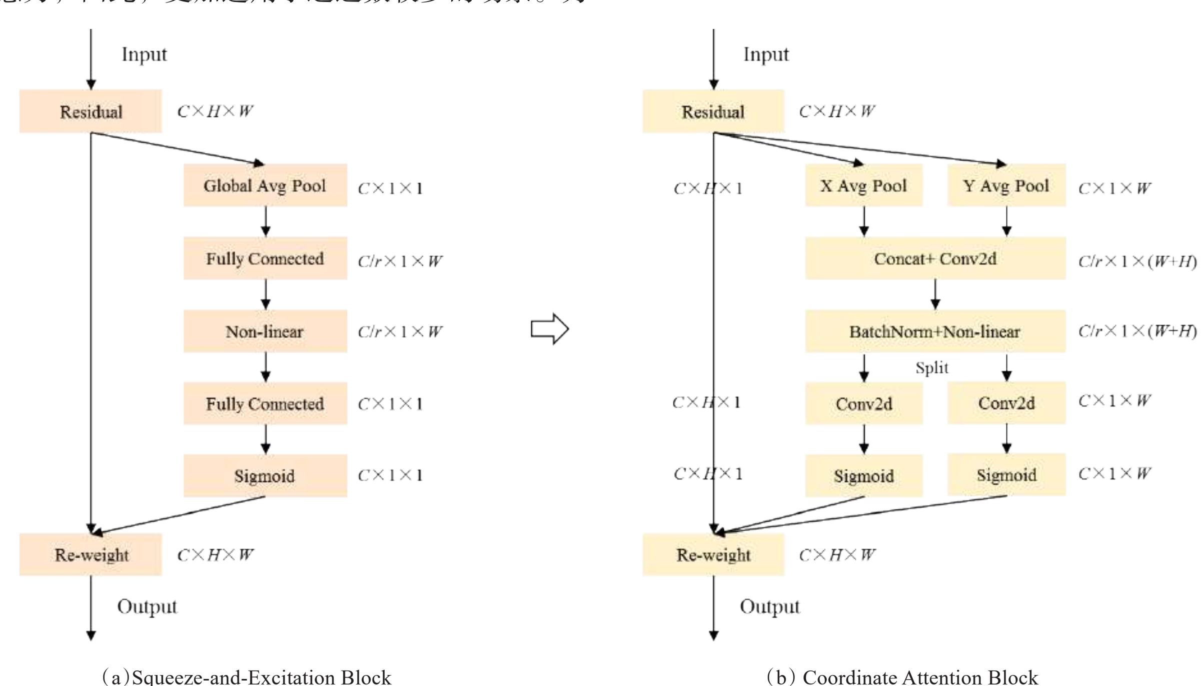


图5　EfficientNetV2-C网络结构图

Fig. 5　EfficientNetV2-C network structure diagram

制增强关键信息的提取能力，如图6（a）所示。然而，SE注意力机制只考虑通过建模通道关系来重新衡量每个通道的重要性，无法捕捉空间维度上的注意力，因此，更加适用于通道数较多的场景。为增强卷积神经网络的特征提取能力，本研究引入可以同时考虑通道维度和空间维度上的CA注意力机制[26]，如图6（b）所示。



（a）Squeeze-and-Excitation Block　　（b）Coordinate Attention Block

图6　注意力机制原理图

Fig. 6　Schematic diagram of the attention mechanism

CA注意力机制是一种基于类型坐标的注意力机制，通过在通道注意力中嵌入类型信息，使模块在不同空间尺度上对特征图进行自适应的调整，从而扩大移动网络参与范围，实现增强移动网络学习特征的表达能力。CA注意力机制具体步骤可分为Coordinate信息嵌入和Coordinate Attention生成。Coordinate信息嵌入为了避免全局池化丢失空间信息，将全局池化分为两个一维全局池化操作，分为编码水平和垂直方向的特征，如公式（1）～（3）所示。

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{1}$$

$$z_c^h(h) = \frac{1}{W}\sum_{0 \leqslant i < W} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{H}\sum_{0 \leqslant j < H} x_c(jwé) \quad (3)$$

其中，z_c表示与第c信道相关联的输出；$z_c^h(h)$表示沿着水平坐标聚合特征；$z_c^w(w)$表示沿着垂直坐标聚合特征；H表示图像高度，pixel；W表示图像宽度，pixel；x表示给定输入；c为通道个数，个。通过上述公式从不同的方向集成特征，输出一对方向可知的特征图，允许attention block捕捉单方向上的长距离关系同时保留另一个方向上的空间信息，帮助网络更准确地定位目标。Coordinate Attention生成的步骤，首先将宽度与高度方向的特征图进行拼接，通过卷积模块降维，随后将数据批量归一化送入Sigmoid激活函数，最终获取特征图f，如公式（4）所示。

$$f = \delta\left(F_x\left(\left[z^h, z^w\right]\right)\right) \quad (4)$$

其中，F_x为卷积变换函数；［z^h，z^w］表示沿空间维度的级联操作；δ为非线性激活函数；$f \in R^{\frac{C}{r} \times (H+W)}$表示在水平方向和垂直方向上编码空间信息的中间特征图。随后分别在水平与垂直方向利用卷积变换函数将特征图f变换为与输入x具有相同通道数的张量，并沿空间维度将f分成$f^h \in R^{\frac{C}{r} \times H}$和$f^w \in R^{\frac{C}{r} \times W}$两个独立的张量输入至sigmoid激活函数，最终输出两个方向的注意力权重，如公式（5）和公式（6）所示。

$$g^h = \sigma\left(F_h\left(f^h\right)\right) \quad (5)$$

$$g^w = \sigma\left(F_w\left(f^w\right)\right) \quad (6)$$

其中，g^h为水平方向权重；g^w为垂直方向权重，最后在原始特征图上通过乘法加权计算，结合两个方向注意权重并扩展可获取Coordinate Attention输出的y，如公式（7）所示。

$$y_c\left(i,j\right) = x_c\left(i,j\right) \times g_c^h(i) \times g_c^w\left(j\right) \quad (7)$$

2.5.2 损失函数

损失函数作为深度学习模型组成的重要模块之一，通常用于衡量模型的预测结果与真实结果之间差异的指标。合适的损失函数将会给模型训练带来积极的效果。在大多数的分类任务中，通常使用交叉熵（Cross Entropy，CE）作为损失函数。它将模型对于正确类别的预测概率视为最大化的目标，同时惩罚模型对于错误类别的预测概率。对于单个样本来说，交叉熵损失的计算如公式（8）所示。

$$L(y,p) = -\sum y_i \times \log(p_i) \quad (8)$$

其中，y_i表示第i个类别是否为真实类别；p_i表示第i个类别的概率；L（y，p）表示交叉熵函数；y表示测量真实概率分布；p表示预测概率分布。在实际应用中，通常存在采集的数据每个类别之间样本数量相差较大。因此，为了使模型更加关注样本较少的类别，需要将不同类别的损失进行加权，如公式（9）所示。

$$L(y,p) = -\sum w_i \times y_i \times \log(p_i) \quad (9)$$

其中，w_i表示第i个类别的权重。权重的选择通常是基于经验或直觉的，而不是根据数据或严格的统计分析确定的，因此可能导致对权重的选择存在主观性，会影响模型的训练效果。为解决不平衡数据集和难易样本的学习，Lin等[27]提出一种基于二分类交叉熵的改进损失函数Focal Loss，通过引入一个动态缩放因子，动态降低训练过程中易区分样本的权重，更多关注难区分的样本，如公式（10）所示。

$$FL(p) = -\alpha(1-p)^{\lambda}\log\left(p\right) \quad (10)$$

其中，α表示类别权重；λ表示难分样本权重；p表示模型给出的正类别预测概率；（$1-p$）表示负类别的概率。Focal Loss能够在不破坏原数据分布的情况下，解决模型因数据分类不均导致的性能问题。但面对特别困难的样本时会极大影响模型的稳定性，其次Focal Loss作为一种静态的损失函数，无法适应不同类别的分布变化。本研究采用Focal Loss的改进损失函数Class-Balanced Focal Loss（CB-Focal Loss），能够根据样本类别的分布变化动态调整每个类别间的权重，使模型学习更加均衡，防止模型过度偏向数量较多的类别，并且拥有更好的泛化性。具体如公式（11）所示。

$$\mathrm{CB}_{\mathrm{focal}} = -\frac{1-\beta}{1-\beta^{n_y}}\sum_{i=1}^{C}(1-p)^{\lambda}\log\left(\mathrm{p}\right) \quad (11)$$

其中，利用$\frac{1-\beta}{1-\beta^{n_y}}$的自适应权重，代替Focal Loss中的$\alpha$权重，实现根据类别分布动态调整权重参数。

2.6 试验设置与评价指标

本研究中所有网络模型均在Linux服务器下搭建，深度学习模型基于PyTorch实现，版本为1.7.1，Python版本为3.8，CUDA 11.0，Python程序的开发环境为VScode，硬件采用NVIDIA RTX A5000显卡在服务器上进行试验，显存大小为16 G，CPU采用Intel（R） Xeon（R） Platinum 8358P CPU@ 2.60GHz，模型推理训练时采用Efficientnet-m.pth预训练权重。模型训练超参数设置如下：受硬件参数限制batch size设置为16，epoch设置为100，初始学期率为0.001，学习率变化策略为等间隔变化，步幅为2，训练阶段使用SGD[28]优化器，动量设置为0.9，权重衰退设置为0.0001。

为检验模型训练的效果是否能够适用于复杂麦田环境下小麦倒伏类型预测，本研究采用精确度（Accuracy）、准确率（Precision）、召回率（Recall）和F_1分数（F_1-Score）评估模型（公式（12）～公式（15））。TP、TN、FP、FN分别为真正类、真负类、假正类、假负类。其中精确度表示模型正确分类的样本数量占总样本数量的比例，精度越高，表示模型分类的准确性越高，F_1分数是准确率和召回率的综合度量，是分类模型常用的综合评价指标，能够避免精确度或召回率出现单一极大值，并在准确率和召回率之间取得平衡。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F_1\text{ - }Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{15}$$

3 结果与讨论

3.1 机器学习分类结果

图7展示了使用4种机器学习分类器对无人机3种不同飞行高度获取的小麦图像倒伏类型进行分类的平均结果。从图7中可知，对比4种模型的分类效果，SVM分类器表现良好，平均准确率高达81.95%，精准度为83.93%，召回率为86.22%，F_1值为84.10%。而决策树对于图像分类任务存在一定限制，特别是当图像具有大量像素和复杂特征，并且这些特征之间存在高度非线性的关系时，决策树很难捕捉这些复杂特征，导致模型分类效果较SVM差，其平均精度仅为79.56%。而针对K最近邻算法和朴素贝叶斯算法，数据的尺寸和分辨率的变化会极大地影响模型的分类效果。当输入的图像具有高维的像素值特征时，可能会导致“维度灾难”问题，因此这两种模型的平均分类精度仅为59.32%。

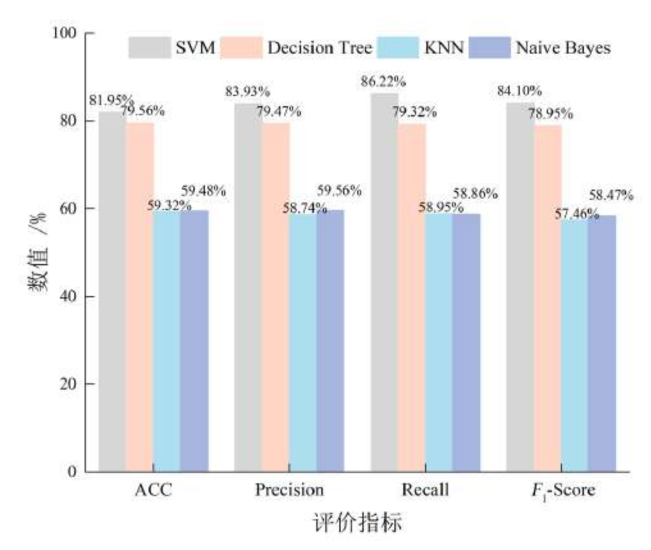


图7 四种机器学习分类器(SVM、DT、KNN、NB)分类结果

Fig. 7 Classification results of four machine learning classifiers (SVM, DT, KNN, NB)

表1展示了SVM分类模型在无人机3种不同飞行高度下获取小麦图像的不同倒伏类型的分类结果。根据表1可知，不同飞行高度对SVM分类的效果影响较小，当高度为45 m时准确率最高，为83.51%。三个飞行高度下倒伏类型在根部的分类效果最好，其中召回率均达到100%。模型均能够正确匹配所有真实的正例。SVM模型的分类结果值彼此相近，说明输入数据类别的不平衡并不会影响模型的分类效果，且SVM分类模型能够有效解决数据高维问题，对中小型数据集拥有较好的分类效果。

3.2 深度学习分类结果

表2展示了3种深度学习模型的分类结果。从表2可以看出，对于ResNet101和EfficientNetV2，随着无人机飞行高度的上升，两种模型的准确率都

表1 使用SVM分类器对3个高度下每个类别小麦倒伏类型分类的结果

Table 1 Results of using SVM classifier to categorize the types of lodging wheat for each category at three heights

高度/m	评价指标	未倒伏/%	根部倒伏/%	茎部倒伏/%
15	Precision	82.13	81.56	79.43
	Recall	83.45	100.00	72.81
	F_1-Score	84.23	84.23	77.79
	Accuracy/%		81.33	
45	Precision	83.56	95.13	85.50
	Recall	85.11	100.00	79.35
	F_1-Score	84.11	98.73	82.44
	Accuracy/%		83.51	
91	Precision	84.47	85.47	78.02
	Recall	73.97	100.00	81.28
	F_1-Score	82.30	82.45	80.60
	Accuracy/%		81.00	

呈下降趋势。ResNet101的准确率由81.57%下降至78.04%；EfficientNetV2的准确率由84.40%下降至81.61%。实验结果表明，两种深度学习模型的分类性能都受无人机的飞行高度影响，主要原因为在相同的参数条件下，飞行高度越低，获取的图像更清晰，图像中富含内容更丰富且特征更明显，因此深度学习模型能够提取更多有用特征。对比上述两种分类模型，EfficientNetV2的分类结果略优于ResNet101，因此本研究选择EfficientNetV2作为基础分类模型。由表2可得知，EfficientNetV2对每一类的分类结果，每一类的Precision、Recall、F_1分数结果数值间相差较大，呈震荡趋势。例如，高度为15 m时未倒伏的小麦的F_1分数高达88.09%，而茎部倒伏的F_1值仅有79.05%。产生上述问题的主要原因是输入分类模型的数据不均衡。深度学习模型在面对不均衡数据集时容易偏向于预测数量较多的类别，而对数量较少的类别在训练过程中学习的类别特征较少，从而降低了预测能力。

表2 使用3种深度学习分类模型对3种高度下每个类别小麦倒伏类型分类结果

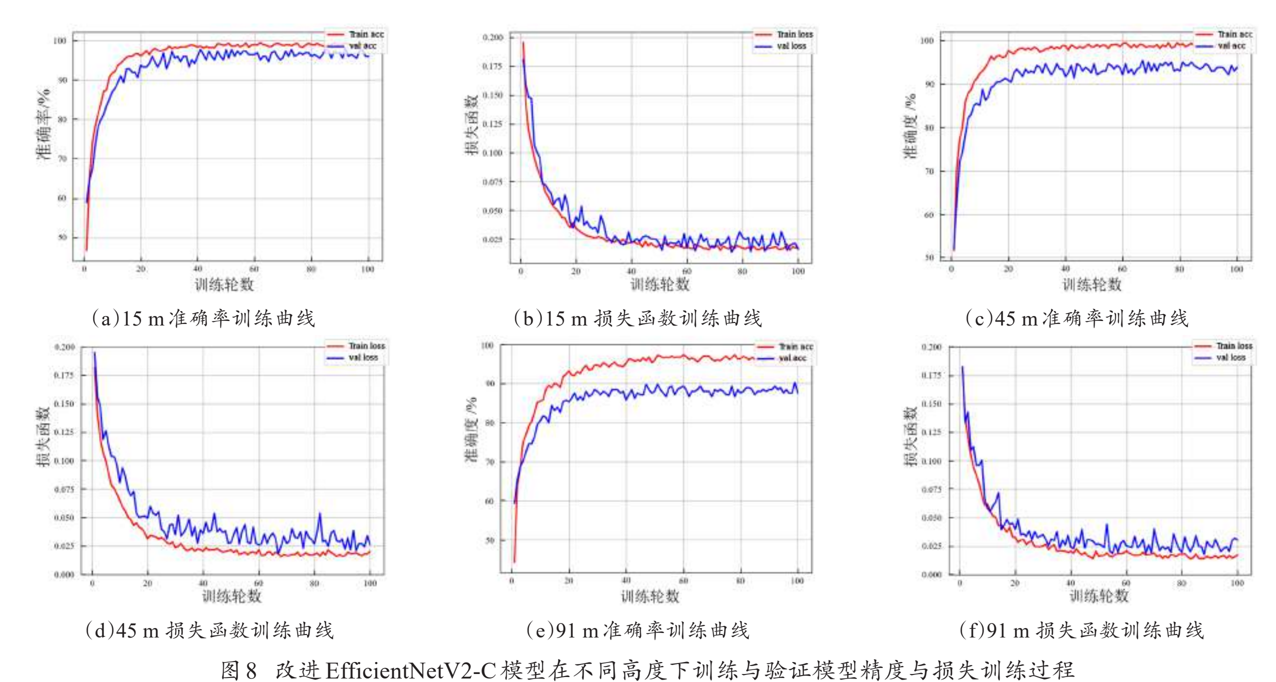
Table 2 Results of using the three deep learning classification models to categorize the types of wheat lodging for each category at the three heights

		ResNet101			EfficientNetV2			EfficientNetV2-C		
高度/m	倒伏类型	Precision/%	Recall/%	F_1-Score/%	Precision/%	Recall/%	F_1-Score/%	Precision/%	Recall/%	F_1-Score/%
15	未倒伏	77.42	90.00	83.24	80.08	92.50	88.09	97.53	98.75	98.14
	根部倒伏	84.71	84.71	84.71	88.59	85.53	87.03	96.59	100.00	98.27
	茎部倒伏	83.12	71.11	76.65	82.22	73.78	79.05	98.84	94.44	96.59
	Accuracy/%		81.57			84.40			97.65	
45	未倒伏	77.08	92.50	84.09	83.72	90.00	86.75	84.62	96.25	90.06
	根部倒伏	84.21	75.29	79.50	79.55	82.35	80.92	92.13	96.47	94.25
	茎部倒伏	77.11	71.11	73.99	76.54	68.89	72.51	93.59	81.11	86.90
	Accuracy/%		79.22			82.00			92.5	
91	未倒伏	79.79	93.75	86.21	81.11	91.25	85.88	87.95	91.25	89.57
	根部倒伏	78.31	76.47	77.38	85.33	75.29	80.00	87.21	93.75	90.36
	茎部倒伏	75.64	65.56	70.24	73.33	73.33	73.33	92.41	81.11	86.39
	Accuracy/%		78.04			81.61			90.59	

对比EfficientNetV2与EfficientNetV2-C，从表2可以得知，虽然EfficientNetV2-C也受到无人机飞行高度的影响，但改进后的模型的分类效果相较于原模型的平均准确率有明显提升。随着飞行高度的升高，准确率由97.65%降低至90.59%。此外，对比不同无人机飞行高度下3种倒伏类型的分类效果，EfficientNetV2-C在3种高度下的每一类分类的Precision、Recall、F_1分数结果数值间相差较小，表明本研究改进的模型能够有效地解决因数据不均衡而导致的模型分类结果不均衡且分类效果差等问题。其次，根部倒伏在不同高度下的分类效果均优于其他两类，这可能是因为根部倒伏的数据质量要

高于其他两类。

用于小麦倒伏类型分类的EfficientNetV2-C精度与损失函数训练过程如图8所示。在结合新的CA注意力机制与CB-Focal Loss的条件下，模型在训练不同高度的数据集时，训练精度与验证精度均保持相同趋势，损失函数训练曲线与验证曲线在保持相同下降趋势的情况下验证集损失函数略大于训练集损失函数，这表明模型在训练时遵循了一个均匀而温和的趋势，且损失函数最终趋于收敛，模型不存在过拟合现象。模型预测的结果如图9所示。



(a)15 m准确率训练曲线　(b)15 m损失函数训练曲线　(c)45 m准确率训练曲线

(d)45 m损失函数训练曲线　(e)91 m准确率训练曲线　(f)91 m损失函数训练曲线

图8　改进EfficientNetV2-C模型在不同高度下训练与验证模型精度与损失训练过程

Fig. 8　Improvement of EfficientNetV2-C model for training and validation of model accuracy and loss training process at different altitudes

3.3　机器学习模型与深度学习模型比较

本节基于3.1节机器学习SVM分类结果与3.2节深度学习EfficientNetV2-C分类结果。基于本研究的数据集，基于深度学习的EfficientNetV2-C模型拥有更好的分类性能，与SVM相比平均准确率提升11.63%，平均精确度提升8.40%，平均召回率提升6.34%，平均 F_1 分数提升8.18%。EfficientNetV2-C在15 m时提升最为明显，其中准确率提升16.32%，精确度提升16.61%，召回率提升12.31%，F_1 分数提升15.59%。综上所述，对比机器学习SVM分类模型，因SVM分类器通常使用手工设计的特征，这些特征可能不足以捕捉到复杂的图像特征，而深度学习模型EfficientNetV2-C在较低高度时提升分类性能明显。这是由于无人机飞行较低高度时拍摄照片包含特征信息更丰富，深度学习模型更适用于高维数据且用于处理具有大量特征的数据，并且深度学习可以学习和表示复杂的非线性关系，从而提高特征提取的表达能力和模型的分类性能。综合对比分类性能和分类结果，本研究认为EfficientNetV2-C是基于无人机图像的小麦倒伏类型分类的最佳模型。

4　结　论

本研究旨在利用现有图像处理技术对无人机在3种不同飞行高度（15、45、91 m）下获取的小麦试验田RGB图像的倒伏类型进行检测。为实现这一目标，本研究利用自主设计的麦田图像自动分割算法获取的单个小麦试验田图像集，使用融合CA注意力机制与Focal Loss损失函数的EfficientNetV2-C改进模型对其进行倒伏类型检测。结果表明，EfficientNetV2-C拥有优异的倒伏类型检测性能，3种高度下的平均准确率均大于90.00%。此外，无人机飞行高度会对深度学习模型倒伏类型检测性能造成影响，具体表现为随着飞行高度的上



注：其中每一小幅图左侧表示长矩形试验田尺寸为1.5 m×15 m的预测结果。右侧表示短矩形试验田尺寸为1.5 m×3.7 m的预测结果

图9 EfficientNetV2-C对不同高度下3种小麦倒伏类型预测

Fig. 9 EfficientNetV2-C prediction of three types of collapse at different heights

升，模型分类准确率随之下降，在最低高度15 m时实现最高精度（97.65%）。与其他模型（SVM、KNN、DT、NB、ResNet101、EfficientNetV2）相比，模型在不同高度下的平均准确率均提升10%以上。然而，本研究仅针对小麦倒伏类型进行研究，因此存在一定的局限性。为进一步验证模型的普适性，未来的研究可以扩大研究种类，将其应用到其他农作物的倒伏检测验证中。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 胡卫国, 曹廷杰, 杨剑, 等. 小麦新品种(系)抗倒性及产量构成因素评价[J]. 种子, 2021, 40(2): 110-115.
HU W G, CAO T J, YANG J, et al. Evaluation of lodging resistance and yield components of new wheat varieties (lines)[J]. Seed, 2021, 40(2): 110-115.

[2] WU W, MA B L. A new method for assessing plant lodging and the impact of management options on lodging in canola crop production[J]. Scientific reports, 2016, 6: ID 31890.

[3] PINTHUS M J. Lodging in wheat, barley, and oats: The phenomenon, its causes, and preventive measures[J]. Advances in agronomy, 1974, 25: 209-263.

[4] 王芬娥, 黄高宝, 郭维俊, 等. 小麦茎秆力学性能与微观结构研究[J]. 农业机械学报, 2009, 40(5): 92-95.
WANG F E, HUANG G B, GUO W J, et al. Mechanical properties and micro-structure of wheat stems[J]. Transactions of the Chinese society for agricultural machinery, 2009, 40(5): 92-95.

[5] BERRY P M, SPINK J. Predicting yield losses caused by lodging in wheat[J]. Field crops research, 2012, 137: 19-26.

[6] BERRY P M, STERLING M, SPINK J H, et al. Understanding and reducing lodging in cereals[M]// Advances in agronomy. Amsterdam: Elsevier, 2004: 217-271.

[7] 孙盈盈, 王超, 王瑞霞, 等. 小麦倒伏原因、机理及其对产量和品质影响研究进展[J]. 农学学报, 2022, 12(3): 1-5.
SUN Y Y, WANG C, WANG R X, et al. Wheat lodging: Cause and mechanism and its effect on wheat yield and quality[J]. Journal of agriculture, 2022, 12(3): 1-5.

[8] 赵静, 闫春雨, 杨东建, 等. 基于无人机多光谱遥感的台风灾后玉米倒伏信息提取[J]. 农业工程学报, 2021, 37(24): 56-64.
ZHAO J, YAN C Y, YANG D J, et al. Extraction of maize lodging information after typhoon based on UAV multispectral remote sensing[J]. Transactions of the Chinese society of agricultural engineering, 2021, 37(24): 56-64.

[9] 董锦绘, 杨小冬, 高林, 等. 基于无人机遥感影像的冬小麦倒伏面积信息提取[J]. 黑龙江农业科学, 2016(10): 147-152.
DONG J H, YANG X D, GAO L, et al. Information extraction of winter wheat lodging area based on UAV remote sensing image[J]. Heilongjiang agricultural sciences, 2016(10): 147-152.

[10] 刘良云, 王纪华, 宋晓宇, 等. 小麦倒伏的光谱特征及遥感监测[J]. 遥感学报, 2005, 9(3): 323-327.
LIU L Y, WANG J H, SONG X Y, et al. The canopy spectral features and remote sensing of wheat lodging[J]. Journal of remote sensing, 2005, 9(3): 323-327.

[11] ZHANG Z, FLORES P, IGATHINATHANE C, et al. Wheat lodging detection from UAS imagery using machine learning algorithms[J]. Remote sensing, 2020, 12(11): ID 1838.

[12] BENDIG J, YU K, AASEN H, et al. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley[J]. International journal of applied earth observation and geoinformation, 2015, 39: 79-87.

[13] DU M M, NOGUCHI N. Multi-temporal monitoring of wheat growth through correlation analysis of satellite images, unmanned aerial vehicle images with ground variable[J]. IFAC-PapersOnLine, 2016, 49(16): 5-9.

[14] LU Y Z, LU R F. Detection of surface and subsurface defects of apples using structured-illumination reflectance imaging with machine learning algorithms[J]. Transactions of the ASABE, 2018, 61(6): 1831-1842.

[15] NAIK D L, KIRAN R. Identification and characterization of fracture in metals using machine learning based texture

recognition algorithms[J]. Engineering fracture mechanics, 2019, 219: ID 106618.
[16] RAJAPAKSA S, ERAMIAN M, DUDDU H, et al. Classification of crop lodging with gray level co-occurrence matrix[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, New Jersey, USA: IEEE, 2018: 251-258.
[17] ZHANG Z, IGATHINATHANE C, FLORES P, et al. UAV mission height effects on wheat lodging ratio detection[M]// Unmanned aerial systems in precision agriculture. Singapore: Springer, 2022: 73-85.
[18] YU J, CHENG T, CAI N, et al. Wheat lodging segmentation based on Lstm_PSPNet deep learning network[J]. Drones, 2023, 7(2): ID 143.
[19] NEUPANE B, HORANONT T, HUNG N D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV) [J]. PLoS one, 2019, 14 (10): ID e0223906.
[20] MAHESH B. Machine learning algorithms: A review[J]. International journal of science and research, 2020, 9(1): 381-386.
[21] 韩安太, 郭小华, 廖忠, 等. 基于压缩感知理论的农业害虫分类方法[J]. 农业工程学报, 2011, 27(6): 203-207.
HAN A T, GUO X H, LIAO Z, et al. Classification of agricultural pests based on compressed sensing theory[J]. Transactions of the Chinese society of agricultural engineering, 2011, 27(6): 203-207.
[22] GUO G, WANG H, BELL D, et al. On the move to meaningful internet systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings[M]. Berlin: Springer Berlin Heidelberg, 2003.
[23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 770-778.
[24] TAN M X, LE Q V. EfficientNetV2: Smaller models and faster training[EB/OL]. arXiv: 2104.00298, 2021
[25] ZHOU D Q, HOU Q B, CHEN Y P, et al. Rethinking bottleneck structure for efficient mobile network design[EB/OL]. arXiv: 2007.02269, 2020.
[26] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2021: 13708-13717.
[27] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2999-3007.
[28] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.

Wheat Lodging Types Detection Based on UAV Image Using Improved EfficientNetV2

LONG Jianing[1,2], ZHANG Zhao[1,2*], LIU Xiaohang[1,2], LI Yunxia[1,2], RUI Zhaoyu[1,2], YU Jiangfan[1,2], ZHANG Man[1,2], FLORES Paulo[3], HAN Zhexiong[4,5], HU Can[6], WANG Xufeng[6]

（1. *College of Information and Telecommunications*, *China Agricultural University*, *Beijing* 100080, *China*; 2. *Key Laboratory of Agricultural Information Acquisition Technology*, *Ministry of Agriculture and Rural Affairs*, *China Agricultural University*, *Beijing* 100083, *China*; 3. *Department of Agricultural and Bioengineering*, *North Dakota State University*, *Fargo* 58102, *USA*; 4. *Department of Biosystems Engineering*, *Kangwon University*, *Chuncheon* 24341, *Korea*; 5. *Department of Smart Agriculture Interdisciplinary*, *Gangwon University*, *Chuncheon* 24341, *Korea*; 6. *College of Mechanical and Electronic Engineering*, *Tarim University*, *Alar*, *Xinjiang* 843300, *China*）

Abstract:

[Objective] Wheat, as one of the major global food crops, plays a key role in food production and food supply. Different influencing factors can lead to different types of wheat lodging, e.g., root lodging may be due to improper use of fertilizers. While stem lodging is mostly due to harsh environments, different types of wheat lodging can have different impacts on yield and quality. The aim of this study was to categorize the types of wheat lodging by unmanned aerial vehicle (UAV) image detection and to investigate the effect of UAV flight altitude on the classification performance.

[Methods] Three UAV flight altitudes (15, 45, and 91 m) were set to acquire images of wheat test fields. The main research methods contained three parts: an automatic segmentation algorithm, wheat classification model selection, and an improved classification mod-

el based on EfficientNetV2-C. In the first part, the automatic segmentation algorithm was used to segment the UAV to acquire the wheat test field at three different heights and made it into the training dataset needed for the classification model. The main steps were first to preprocess the original wheat test field images acquired by the UAV through scaling, skew correction, and other methods to save computation time and improve segmentation accuracy. Subsequently, the pre-processed image information was analyzed, and the green part of the image was extracted using the super green algorithm, which was binarized and combined with the edge contour extraction algorithm to remove the redundant part of the image to extract the region of interest, so that the image was segmented for the first time. Finally, the idea of accumulating pixels to find sudden value added was used to find the segmentation coordinates of two different sizes of wheat test field in the image, and the region of interest of the wheat test field was segmented into a long rectangle and a short rectangle test field twice, so as to obtain the structural parameters of different sizes of wheat test field and then to generate the dataset of different heights. In the second part, four machine learning classification models of support vector machine (SVM), K nearest neighbor (KNN), decision tree (DT), and naive bayes (NB), and two deep learning classification models (ResNet101 and EfficientNetV2) were selected. Under the unimproved condition, six classification models were utilized to classify the images collected from three UAVs at different flight altitudes, respectively, and the optimal classification model was selected for improvement. In the third part, an improved model, EfficientNetV2-C, with EfficientNetV2 as the base model, was proposed to classify and recognized the lodging type of wheat in test field images. The main improvement points were attention mechanism improvement and loss function improvement. The attention mechanism was to replace the original model squeeze and excitation (SE) with coordinate attention (CA), which was able to embed the position information into the channel attention, aggregate the features along the width and height directions, respectively, during feature extraction, and capture the long-distance correlation in the width direction while retaining the long-distance correlation in the length direction, accurate location information, enhancing the feature extraction capability of the network in space. The loss function was replaced by class-balanced focal loss (CB-Focal Loss), which could assign different loss weights according to the number of valid samples in each class when targeting unbalanced datasets, effectively solving the impact of data imbalance on the classification accuracy of the model.

[Results and Discussions] Four machine learning classification results: SVM average classification accuracy was 81.95%, DT average classification accuracy was 79.56%, KNN average classification accuracy was 59.32%, and NB average classification accuracy was 59.48%. The average classification accuracy of the two deep learning models, ResNet101 and EfficientNetV2, was 78.04%, and the average classification accuracy of ResNet101 was 81.61%. Comparing the above six classification models, the EfficientNetV2 classification model performed optimally at all heights. And the improved EfficientNetV2-C had an average accuracy of 90.59%, which was 8.98% higher compared to the average accuracy of EfficientNetV2. The SVM classification accuracies of UAVs at three flight altitudes of 15, 45, and 91 m were 81.33%, 83.57%, and 81.00%, respectively, in which the accuracy was the highest when the altitude was 45 m, and the classification results of the SVM model values were similar to each other, which indicated that the imbalance of the input data categories would not affect the model's classification effect, and the SVM classification model was able to solve the problem of high dimensionality of the data efficiently and had a good performance for small and medium-sized data sets. The SVM classification model could effectively solve the problem of the high dimensionality of data and had a better classification effect on small and medium-sized datasets. For the deep learning classification model, however, as the flight altitude increases from 15 to 91 m, the classification performance of the deep learning model decreased due to the loss of image feature information. Among them, the classification accuracy of ResNet101 decreased from 81.57% to 78.04%, the classification accuracy of EfficientNetV2 decreased from 84.40% to 81.61%, and the classification accuracy of EfficientNetV2-C decreased from 97.65% to 90.59%. The classification accuracy of EfficientNetV2-C at each of the three altitudes. The difference between the values of precision, recall, and F_1-Score results of classification was small, which indicated that the improved model in this study could effectively solve the problems of unbalanced model classification results and poor classification effect caused by data imbalance.

[Conclusions] The improved EfficientNetV2-C achieved high accuracy in wheat lodging type detection, which provides a new solution for wheat lodging early warning and crop management and is of great significance for improving wheat production efficiency and sustainable agricultural development.

Key words: wheat lodging types; image processing; deep learning; unbalanced data; machine learning; UAV

Foundation items: National Key Research and Development Program of China (2022YFD2001500)

(登陆www.smartag.net.cn免费获取电子版全文)